EDITORIAL

Digital behavioural data should inform mental health screening

Sean D. Young

BJPsych Bulletin (2026) Page 1 of 3, doi:10.1192/bjb.2026.10320

Department of Emergency Medicine, School of Medicine, University of California, Irvine, California, USA

Correspondence to Sean D. Young (syoung5@hs.uci.edu)

First received 9 Apr 2026 UTC, final revision 6 Sep 2026 UTC, accepted 28 Sep 2026 UTC

© The Author(s), 2026. Published by Cambridge University Press on behalf of Royal College of Psychiatrists. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Summary Digital behavioural data from smartphones, wearables and artificial intelligence-enabled systems are increasingly available alongside mental health screening questionnaires. However, there is limited guidance on when such signals are appropriate for mental health screening contexts and how they should be interpreted without overreach into diagnosis. We outline four principles (detectable change, context, informing timing and sustained utility) to guide how digital data can inform interpretation and follow-up while preserving the role of validated screening instruments.

Keywords Digital phenotyping; mental health screening; PHQ-9; digital biomarkers.

Widely used mental health screening instruments such as the Patient Health Questionnaire (PHQ) and the Generalized Anxiety Disorder scale (GAD) provide brief, standardised assessments of symptom severity and are deeply embedded in clinical care and research.[1,2] Although efficient and interpretable, these tools were designed primarily for episodic trait-like summaries of symptoms over a defined period, with less capacity to directly capture how functioning and behaviour change between screening points.[3,4]

Clinical practice and research contexts have evolved. Remote care, online longitudinal studies and widespread use of consumer digital devices have made digital behavioural data available between visits, often alongside repeated questionnaire administration.[5–7]

In this article, digital behavioural data refers to measures of behaviour collected through digital technologies, including sleep, physical activity, mobility, communication patterns and technology use. These measures are typically collected through smartphones, wearable devices or other connected technologies and can provide information to clinicians about day-to-day change that may differ from retrospective self-report.[8,9]

As a result of the wide use of these devices and their data, mental health screening scores are increasingly interpreted in the presence of digital behavioural data, yet there is no shared clinical framework for deciding when or which digital signals are suitable for screening contexts, how they should be interpreted alongside questionnaires and how to avoid extending them into diagnostic roles. Clinicians and researchers are increasingly confronted with the challenge of access to multiple sources of information, including digital signals, but without clear guidance on how those sources should be integrated.[8]

Using digital approaches to support screening: a decision framework

I propose a framework in which validated questionnaires remain the primary standardised instruments, while digital tools and their data inform clinical attention and longitudinal interpretation. Under this approach, changes in screening scores are interpreted alongside observable changes in daily functioning rather than treated as stand-alone indicators of improvement or deterioration. I outline four principles for evaluating when and how digital tools can support screening and longitudinal interpretation, including how data should be interpreted, when they add value and how they relate to established instruments. The principles outlined here apply across both primary care and specialty settings.

Principle 1: Detectable change

Digital data must reliably show that something in a person's behaviour or routine has changed over time. Questionnaires capture self-reported symptoms at discrete time points, but often miss changes in daily functioning between visits.[10] For digital data to contribute to screening beyond questionnaires or patient interviews, clinicians must be able to describe in simple terms what kind of change occurred. For example, changes in mobility may reflect shifts in daily routine or social activity.[11,12] Evidence that a patient slept three fewer

https://doi.org/10.1192/bjb.2026.10320 Published online by Cambridge University Press

hours per night over the past week relative to their recent baseline provides a concrete, functionally interpretable change that can be discussed clinically. Changes in phone use or communication patterns may reflect changes in engagement or cognitive effort. These signals do not indicate diagnosis or treatment need, but rather that 'something is different from usual'. If a digital signal cannot be translated into a clinically interpretable behavioural change, it is not appropriate for use in a screening context. Highly multivariate model outputs that lack clear links to behaviour may be difficult to interpret meaningfully in screening contexts, even when statistically predictive.

Principle 2: Context

Digital data must provide context that helps interpret questionnaire responses. Questionnaires are effective at capturing symptom severity at a single point in time,[1,2] but often lack context about what is happening in a person's daily functioning between assessments.

Digital data can provide context that helps situate questionnaire scores within patterns of everyday behaviour over time. For example, a person may report feeling less depressed on a questionnaire, but their sleep, activity or daily routines may remain highly disrupted. Conversely, a person may report stable symptoms while daily functioning steadily improves. In these cases, digital data do not contradict questionnaire results, but help clarify how results should be interpreted.

Consider a patient with stable PHQ scores across two visits. In the intervening month, passive sleep data show a sustained reduction in total sleep time and greater night-to-night variability relative to that patient's baseline. Under this framework, the sleep change does not diagnose depression or determine treatment but provides interpretable context that prompts follow-up, whether the patient is experiencing insomnia, medication effects, shift-work changes, caregiving demands or worsening anxiety. The screening tool remains the standard measure, and the discrepancy provides a reason to ask whether stability in the score also reflects stability in functioning.

When digital and self-reported data disagree

Digital data and self-report will not always converge, and disagreement should not be treated as evidence that one source is necessarily correct. Discordance may reflect limitations in sensor accuracy, differences between what a device measures and what a patient experiences, contextual factors, missing data or inaccurate self-report. For example, wearable data suggesting apparently normal sleep would not by itself invalidate a patient's report of severe sleep disturbance; the device and the patient may be measuring different aspects of sleep.

Unexplained discordance should therefore prompt further clinical inquiry rather than automatic preference for either source. Clinicians should consider whether the measures assess the same construct, whether the digital measure has adequate validity for the proposed interpretation and whether contextual factors explain the difference. The framework is intended to support clinical screening, not forensic evaluation. When digital data require the patient's permission to access, declining to share them should not itself be interpreted as evidence about symptoms or credibility.

Principle 3: Informing timing

Digital data must help inform when screening or follow-up should occur and what questions to ask next. Questionnaires are typically administered at fixed times or in response to patient report, which means screening may be delayed or missed when people do not openly describe emotional distress.[13–15] This principle is most relevant when symptoms are underreported or fluctuate between visits.

The role of digital data is not to determine diagnosis, but to guide attention within existing screening workflows. For example, sustained changes in sleep, activity or engagement may suggest that screening should be initiated or repeated, even if a person has not reported feeling worse. In this role, digital signals should trigger attention and follow-up, not serve as independent screening scores or diagnostic proxies.

Digital signals can also help guide follow-up after screening. A stable questionnaire score may warrant different questions, depending on whether daily functioning is improving or deteriorating. In this way, digital data help clinicians decide when to look more closely and where to focus inquiry, rather than acting as screening tools themselves.

This principle also applies to artificial intelligence-based conversational interviewing systems (AICIS), that support screening by clarifying symptom time course, functional impact or ambiguity in self-reported responses.[16] For example, AICIS may be used to ask follow-up questions when responses are unclear, probe functional consequences when questionnaire scores change or adjust question sequencing using conversational logic rather than a fixed item order. After someone confirms symptoms for a PHQ item, the software can ask a short follow-up question and summarise the answers for the clinician. In this way, AICIS can refine how information is elicited while remaining anchored to existing screening constructs, thresholds and diagnostic frameworks.

Principle 4: Sustained utility

Digital data must remain meaningful over time. Questionnaires rely on stable wording and scoring, but digital data are collected in environments where behaviour, technology and social norms change. As a result, a digital signal that appears useful at one point in time may stop meaning what we think it means in the future.

Digital data used to support mental health screening must therefore remain interpretable as behaviour and context evolve. For example, reduced geographical mobility might once have strongly reflected social withdrawal or functional impairment, but may now also reflect remote working, caregiving responsibilities or changes in how daily activities are organised. In this case, similar behavioural patterns may correspond to different clinical contexts than they did previously.

https://doi.org/10.1192/bjb.2026.10320 Published online by Cambridge University Press

Even when digital data provide detectable change, context or guidance about when to screen, their meaning should not be assumed to remain stable. Changes in devices, operating systems, user behaviour, social norms or data collection methods should prompt periodic reassessment of whether a signal continues to inform the same screening decisions.

Clinical implications

Digital data should not replace mental health screening tools or function as alternative diagnostic measures. Instead, they can provide additional context that helps clinicians interpret screening results more carefully. For example, digital signals may indicate whether a change in a screening score is accompanied by a change in daily functioning, whether screening should be repeated or which follow-up questions are most relevant when responses are ambiguous. When digital and self-reported information conflict, the discrepancy itself should prompt clarification rather than automatic preference for either source.

Without such context, changes in screening scores may be overinterpreted as clinical improvement or deterioration when they instead reflect reporting variability, situational stressors or measurement noise. Digital data can therefore help clinicians identify when screening results warrant closer inquiry, without functioning as independent evidence of diagnosis, treatment response or patient credibility. Implementation should incorporate known limitations of digital devices, including privacy concerns, the need to include atypical clinical presentations in training data and training on a diverse sample.[9,17]

Extending mental health screening beyond traditional self-report instruments does not require replacing standard screening tools or shifting diagnostic responsibility to digital data sources. Rather, it requires clear criteria for how additional information is allowed to influence screening decisions. By specifying when digital data indicate meaningful change, provide useful context, inform the timing of screening and follow-up, and remain interpretable over time, the framework presented here clarifies how digital approaches can support screening while preserving interpretability, feasibility and equity in clinical care. Although illustrated here in mental health screening, these principles may also apply to other domains of digital medicine in which continuous behavioural or physiological data are interpreted alongside episodic clinical measures.

About the author

Sean D. Young is a professor in the Department of Emergency Medicine, School of Medicine, University of California, Irvine, California, USA.

Acknowledgements

I thank Arjuna Ugarte for feedback on a prior version of the manuscript. The author made use of ChatGPT version 5.5 to copyedit this manuscript in 2026. It was used with modifications.

Funding

S.D.Y. received funding from the National Institute on Minority Health and Health Disparities and the National Center for Complementary and Integrative Health (grant numbers: R01MD018548, 5 R01 MD019765, UG3 AI176592).

Declaration of interest

S.D.Y. reports consulting related to digital mental health, including advisory work with Sonar Mental Health.

References

1 Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16**: 606–13.

2 Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006; **166**: 1092–7.

3 Dumenci L, Kroenke K, Keefe FJ, Ang DC, Slover J, Perera RA, et al. Disentangling trait versus state characteristics of the Pain Catastrophizing Scale and the PHQ Depression Scale. *Eur J Pain* 2020; **24**: 1624–34.

4 Young SD, Cumberland WG, Fehmie DA, Ugarte D. Instructional emphasis and PHQ-9 and GAD-7 scores among participants with anxiety or depression: a randomized clinical trial. *JAMA Netw Open* 2026; **9**: e2623952.

5 Tan SY, Sumner J, Wang Y, Wenjun Yip A. A systematic review of the impacts of remote patient monitoring (RPM) interventions on safety, adherence, quality-of-life and cost-related outcomes. *NPJ Digit Med* 2024; **7**: 192.

6 Cho PJ, Olaye IM, Shandhi MMH, Daza EJ, Foschini L, Dunn JP. Identification of key factors related to digital health observational study adherence and retention by data-driven approaches: an exploratory secondary analysis of two prospective longitudinal studies. *Lancet Digit Health* 2025; **7**: e23–34.

7 Young SD, Schneider J. Clinical care, research, and telehealth services in the era of social distancing to mitigate COVID-19. *AIDS Behav* 2020; **24**: 2000–2.

8 Torous J, Linardon J, Goldberg SB, Sun S, Bell I, Nicholas J, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. *World Psychiatry* 2025; **24**: 156–74.

9 Reinertsen E, Clifford GD. A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. *Physiol Meas* 2018; **39**: 05TR01.

10 Jung HW, Kim DY, Lee I, Kim O, Lee S, Lee S, et al. Key features of digital phenotyping for monitoring mental disorders: systematic review. *J Med Internet Res* 2025; **27**: e77331.

11 Garett RR, Yang J, Zhang Q, Young SD. An online advertising intervention to increase adherence to stay-at-home-orders during the COVID-19 pandemic: an efficacy trial monitoring individual-level mobility data. *Int J Appl Earth Obs Geoinf* 2022; **108**: 102752.

12 Hassani M, De Haro C, Flores L, Emish M, Kim S, Kelani Z, et al. Exploring mobility data for enhancing HIV care engagement in Black/African American and Hispanic/Latinx individuals: a longitudinal observational study protocol. *BMJ Open* 2023; **13**: e079900.

13 Ford J, Thomas F, Byng R, McCabe R. Use of the Patient Health Questionnaire (PHQ-9) in practice: interactions between patients and physicians. *Qual Health Res* 2020; **30**: 2146–59.

14 Eack SM, Greeno CG, Lee BJ. Limitations of the Patient Health Questionnaire in identifying anxiety and depression: many cases are undetected. *Res Soc Work Pract* 2006; **16**: 625–31.

15 Collins RE, Lopez LM, Marteau TM. Emotional impact of screening: a systematic review and meta-analysis. *BMC Public Health* 2011; **11**: 603.

16 Sikström S, Boehme RA, Mirström M, Agbotsoka T, Győri G, Lasota M, et al. Generative AI-assisted clinical interviewing of mental health. *Sci Rep* 2025; **15**: 37737.

17 Young SD. Artificial intelligence in HIV research, policy, and clinical care. *Lancet HIV* 2025; **13**: e138–42.

https://doi.org/10.1192/bjb.2026.10320 Published online by Cambridge University Press